Bioinformatic studies to predict microRNAs with the potential of uncoupling RECK expression from epithelial–mesenchymal transition in cancer cells

AUTHOR(S):

Wang, Zhipeng; Murakami, Ryusuke; Yuki, Kanako; Yoshida, Yoko; Noda, Makoto

CITATION:

Wang, Zhipeng ...[et al]. Bioinformatic studies to predict microRNAs with the potential of uncoupling RECK expression from epithelial–mesenchymal transition in cancer cells. Cancer Informatics 2016, 15: 91-102

ISSUE DATE:

2016-05-19

URL:

http://hdl.handle.net/2433/226663

Bioinformatic Studies to Predict MicroRNAs with the Potential of Uncoupling RECK Expression from Epithelial–Mesenchymal Transition in Cancer Cells

Zhipeng Wang[1], Ryusuke Murakami[2], Kanako Yuki[1], Yoko Yoshida[1] and Makoto Noda[1,3]

[1]Laboratory for Malignancy Control Research, Medical Innovation Center, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [2]Department of Gynecology and Obstetrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [3]Department of Molecular Oncology, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

ABSTRACT: RECK is downregulated in many tumors, and forced *RECK* expression in tumor cells often results in suppression of malignant phenotypes. Recent findings suggest that RECK is upregulated after epithelial-mesenchymal transition (EMT) in normal epithelium-derived cells but not in cancer cells. Since several microRNAs (miRs) are known to target *RECK* mRNA, we hypothesized that certain miR(s) may be involved in this suppression of RECK upregulation after EMT in cancer cells. To test this hypothesis, we used three approaches: (1) text mining to find miRs relevant to EMT in cancer cells, (2) predicting miR targets using four algorithms, and (3) comparing miR-seq data and *RECK* mRNA data using a novel non-parametric method. These approaches identified the miR-183-96-182 cluster as a strong candidate. We also looked for transcription factors and signaling molecules that may promote cancer EMT, miR-183-96-182 upregulation, and RECK downregulation. Here we describe our methods, findings, and a testable hypothesis on how RECK expression could be regulated in cancer cells after EMT.

KEYWORDS: cancer, EMT, RECK, paired data correlation, miR-183-96-182, SOX2, TEAD4

CITATION: Wang et al. Bioinformatic Studies to Predict MicroRNAs with the Potential of Uncoupling RECK Expression from Epithelial–Mesenchymal Transition in Cancer Cells. *Cancer Informatics* 2016:15 91–102 doi: 10.4137/CIN.S34141.

TYPE: Original Research

RECEIVED: September 04, 2015. **RESUBMITTED:** February 24, 2016. **ACCEPTED FOR PUBLICATION:** March 07, 2016.

ACADEMIC EDITOR: J. T. Efird, Editor in Chief

PEER REVIEW: Four peer reviewers contributed to the peer review report. Reviewers' reports totaled 1,601 words, excluding any confidential comments to the academic editor.

FUNDING: Authors disclose no external funding sources.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: mnoda@virus.kyoto-u.ac.jp; wangz@kuhp.kyoto-u.ac.jp

Introduction

RECK is downregulated in many types of tumors. Forced expression of RECK inhibits tumor angiogenesis, invasion, metastasis, and/or proliferation, depending on the cell system.[1–5] However, the molecular mechanisms of RECK downregulation in tumor cells remain largely unknown.

The epithelial–mesenchymal transition (EMT) is a process by which epithelial cells lose their apicobasal polarity and cell–cell adhesion and gain migratory and invasive properties, resembling those of mesenchymal cells. EMT is essential for numerous developmental processes, including mesoderm formation and neural tube formation. EMT has also been shown to occur during wound healing, organ fibrosis, cancer invasion, and metastasis.[6,7]

Accumulating evidence indicates the involvement of a number of microRNAs (miRs) in the process of carcinogenesis, as well as cancer progression.[8] Based on their activities, cancer-associated miRs can be divided into two groups: oncomiRs and tumor suppressor miRs. A single miR may target multiple mRNAs, while one mRNA may be targeted by multiple miRs, providing a basis for an intricate network of gene regulation. Hence, bioinformatic analyses and evaluation of large sets of data and publications already accumulated in this field of studies should be a reasonable approach for deducing the cellular functions of miRs. For example, numerous papers on miRs relevant to EMT as well as miRs induced by various stimuli (such as hypoxia and growth factors) and targeting *RECK* mRNA have already been published (see below).

We previously found that TGFβ-induced EMT was accompanied by RECK upregulation in nontumorigenic epithelial cell lines (MCF10A and HMLE), but not in carcinoma-derived cell lines (MCF7 and A549).[9] *RECK* overexpression did not affect the process of EMT but negatively regulated cell proliferation and migration. Although the exact mechanisms by which RECK expression is uncoupled from EMT in cancer cells remain to be elucidated, one obvious possibility is transcriptional repression of *RECK* gene in cancer cells. However, we found some discrepancy between the levels of *RECK* mRNA and RECK protein in cancer cells and, therefore, speculated whether some cancer-associated miRs might also play roles in this uncoupling. To address this question in this study, we first attempted to find candidate miRs using three approaches *in silico*: (1) text mining to extract

miRs enriched in studies of cancer EMT, (2) prediction of miRs targeting *RECK* mRNA, and (3) analysis of TCGA breast cancer miR-seq and *RECK* mRNA data using a newly developed nonparametric correlation test. These approaches point to the involvement of the miR-183-96-182 cluster in the uncoupling of RECK expression from EMT in cancer cells. We also searched for candidate transcription factors involved in this event using ENCODE, transcription factor ChIP-seq data, ONCOMINE gene expression database, and expression datasets deposited in NCBI GEO. We propose a testable hypothesis based on these findings.

Methods

Collecting relevant abstracts from PubMed. The following sets of key words were used to collect relevant abstracts of original papers from PubMed: for EMT-associated miRs in noncancerous cells, (microRNA[Title/Abstract] OR miRNA[Title/Abstract] OR miR[Title/Abstract]) AND (EMT[Title/Abstract] OR "epithelial-mesenchymal transition"[Title/Abstract] OR "epithelial-to-mesenchymal"[Title/Abstract]) NOT (cancer[Title/Abstract] OR metastasis[Title/Abstract] OR carcinoma[Title/Abstract] OR sarcoma[Title/Abstract] OR tumor[Title/Abstract] OR "review"[Publication Type]) and for EMT-associated miRs in cancer, (microRNA[Title/Abstract] OR miRNA[Title/Abstract] OR miR[Title/Abstract]) AND (EMT[Title/Abstract] OR "epithelial-mesenchymal transition"[Title/Abstract] OR "epithelial-to-mesenchymal"[Title/Abstract]) AND (cancer[Title/Abstract] OR metastasis[Title/Abstract] OR carcinoma[Title/Abstract] OR sarcoma[Title/Abstract] OR tumor[Title/Abstract]) NOT "review"[Publication Type]. The search was performed on November 2, 2015. The abstracts were downloaded as .txt files.

Text mining for miRs with differences between noncancerous and cancer cells. We used R package "pubmed.mineR"[10] to process the abstracts from PubMed. This provided a correspondence table "HGNCdata" that includes approved symbol, approved name, gene synonyms, and so on for genes, but not for miRs. Therefore, we acquired miR-related information from the HUGO Gene Nomenclature Committee (HGNC) website.[11] In the abstracts, several alias/synonyms are used to describe the same miR. Thus, we first mapped the prefix "miR-", "microRNA-", "MicroRNA-", "hsa-miR-", and "mmu-miR-" to the same character "MIR", converting the alias in the abstract to the approved symbol, according to HGNC. The converted abstracts were analyzed using pubmed.mineR; we first used the gene_atomization function to extract the miRs mentioned in the abstracts and then used the searchabsT function to count the abstracts that referred to each miR. For each miR referred, we calculated its proportion to all papers describing cancer EMT or noncancer EMT and tested the null hypothesis that the proportion in cancer EMT = proportion in noncancer EMT using two-tailed prop.test.

Prediction of miR targets. We used four commonly used tools for predicting miRs: miRanda (August 2010 release),[12,13] PicTar,[14] TargetScan,[15–17] and MicroT-CDS (microT v4).[18] For miRanda, we used human target site predictions with good mirSVR score and conserved miR. PicTar predictions in vertebrates were used with the default setting to search for all miRs potentially targeting *RECK*, and to evaluate them based on the PicTar score. TargetScan was used with the default setting, and the Context score and P_{CT} (probability of conserved targeting) were used to evaluate the probability of a miR to bind *RECK* 3'-UTR. For microT-CDS, we used the gene symbol "RECK", gene ID "ENSG00000122707", species "Homo sapiens", and Ensembl version 69 from homologous genes. The results were evaluated using the miTG score.

Data collection from TCGA. We acquired the breast invasive carcinoma dataset from the TCGA website[19] on May 3, 2013. From the total 905 breast cancer cases, we extracted 53 cases for which mRNA and miR data on both cancer tissues and matched normal breast tissues were available. In the TCGA project, mRNA expression data had been acquired using Agilent 244K Custom Gene Expression G4502A-07 platform (alias: AgilentG4502A_07). miR-seq data had been obtained using Illumina Genome Analyzer (alias: illuminaga) system or Illumina Hiseq (alias: illuminahiseq) system. For miR-seq analysis, we used the level 3 data (ie, expression calls for miRs per sample). In the case of mRNA, we used level 2 data (ie, normalized signals per probe) and averaged the expression calls of the probes that correspond to the same gene to obtain a single value for each gene. Agilent G4502A_07 contains five probes for *RECK*. From them, we chose to use two (A_23_P83028 and A_23_P83030) that were corresponding to the 3'-UTR of the full-length *RECK* transcript; the other three probes were ignored in this study, since they may also pick up smaller transcripts whose functions are presently unclear (Supplementary Fig. 1). To avoid division by zero in calculation, we assigned the uniform value of 0.1 to all cases in which the miR expression was undetectable.

The paired data correlation test. To facilitate evaluation of the trends in expression changes, we developed a nonparametric method as follows: by setting appropriate cutoff values, the expression change between a pair of measurements was categorized into three groups (up, down, and no change) and labeled accordingly (1, –1, and 0, respectively). We call this label "Value of Change". To compare the changes in two factors (eg, a miR vs. *RECK* mRNA), the absolute value of the difference between two Values of Changes was used for evaluation; we call this parameter "diff". Thus, diff = 0 indicates two factors changing toward the same direction, diff = 1 indicates no change in one factor and up or down in the other, and diff = 2 indicates two factors changing toward the opposite directions. To evaluate the trends of changes in multiple pairs of data (eg, multiple patients), the sum of diff (termed "diff.sum" score) was used. In this case, a smaller diff.sum score indicates a trend of two factors changing toward the same direction, whereas a larger diff.sum

score indicates a trend of two factors changing toward opposite directions. The *P*-value for each diff.sum score was calculated by simulation to evaluate how likely the diff.sum score can be obtained by chance. Specifically, we first got a vector containing 53 diff values for *RECK* mRNA and a 1046 × 53 matrix containing diff values for 1046 miRs in 53 patients. We then simulated a 1000 times larger matrix, 1,046,000 × 53, by randomly shuffling the position of patients in the original diff matrix for miRs, while keeping the *RECK* diff vector unchanged. Using the *RECK* diff vector and the simulated diff matrix for miRs, we obtained 1,046,000 diff.sum scores. Probabilities for each specific diff.sum were then calculated based on the distribution of the 1,046,000 diff.sum scores. The concept of this nonparametric correlation test for paired data, which we termed the paired data correlation (PDC) test, is illustrated in Figure 1. Based on the distribution of *RECK* mRNA and miRs in cancer and normal breast tissues, we chose cutoff values of 0.8 and 1.25 (±0.3219 in $\log_2$), respectively, for the downregulation and upregulation of *RECK* mRNA, respectively, and 0.5 and 2.0 (±1 in $\log_2$) for those of miRs.

Analysis of the miR-183-96-182 promoter. UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly[20] was used to search for possible transcription factors of the miR-183-96-182 cluster. Based on the tracks related to regulation (eg, Integrated Regulation from ENCODE Tracks)[21] and the position of transcription start site (TSS) of the miR-183-96-183 cluster, we decided to focus on the region up to 16 kb upstream of the miR-183 precursor. Transcription factors suggested to bind this region by the Transcription Factor ChIP-seq (161 factors) from ENCODE with Factorbook Motifs were extracted manually.

Survey of gene expression profile data. We screened gene expression profile data from knockdown, knockout, overexpression, or inhibitor treatment experiments relevant to candidate transcription factors. The NCBI GEO datasets that include information on miR-183-96-182 cluster were based mainly on three platforms, namely, Illumina HumanHT-12 V4.0 expression beadchip (GPL10558), Affymetrix Mouse Gene 1.0 ST Array (GPL6246), and Affymetrix Human Gene 1.0 ST Array (GPL6244). These data were acquired

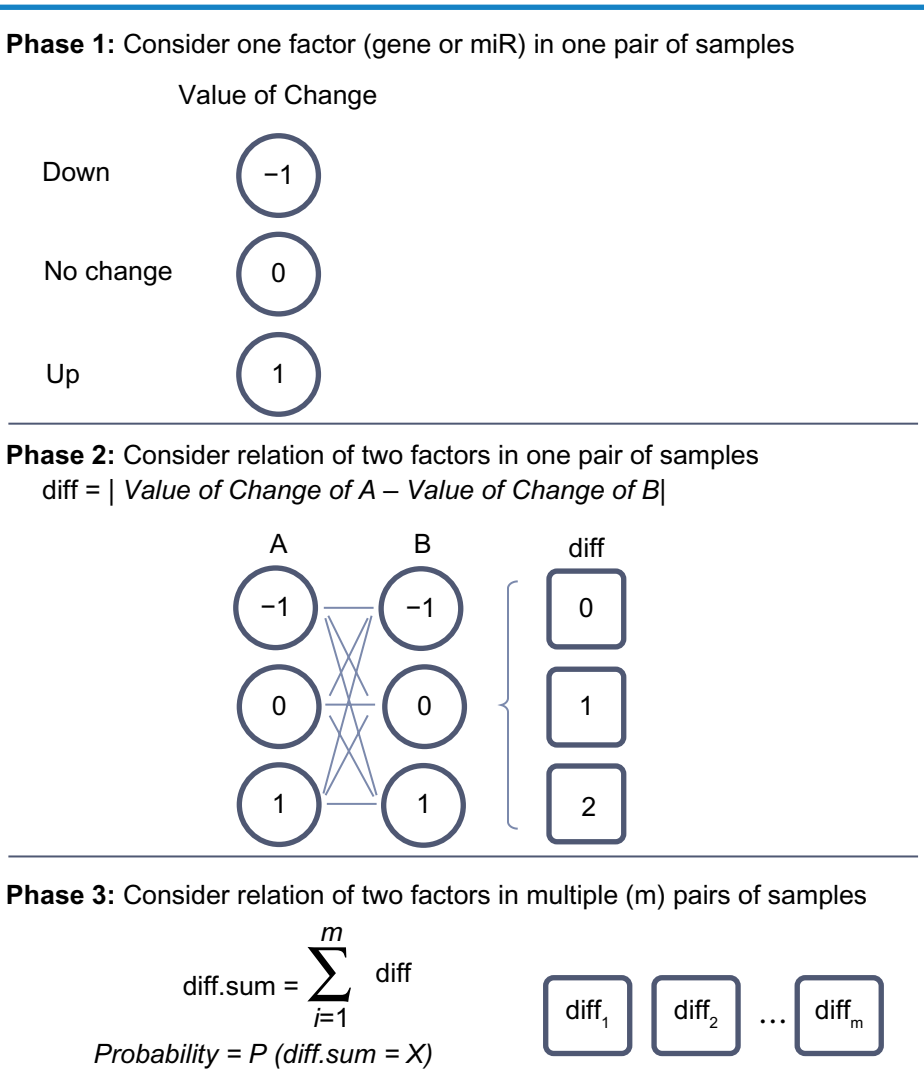

Figure 1. Algorithm of paired data correlation (PDC) test.

using R package "GEOquery",[22] and the values for multiple probes corresponding to the same gene ID were averaged to obtain a representative value for each gene. For important candidates, we extended our search to other array-based platforms as well as RNA-seq data. We also used ONCOMINE 4.5[23] to find the trends in expression of candidate transcription factors for miR-183-96-182 in various cancers. The thresholds used were *P*-value = 0.001, fold change = 2, and gene rank = top 10%.

Results

Text mining for papers describing miRs involved in cancer EMT and noncancer EMT. To examine our hypothesis that miRs may play roles in uncoupling RECK expression from EMT in cancer, we first surveyed original papers in PubMed describing miRs involved in cancer EMT and noncancer EMT using the key word sets described in the "Methods" section. This literature search yielded 887 papers on cancer EMT and 145 papers on noncancer EMT. The most reported miRs in cancer EMT were members of the miR-200 family: miR-200c (121 papers), miR-200b (75 papers), and miR-200a (59 papers). The most reported miR in noncancer EMT was miR-21 (20 papers), which has also been associated with cancer EMT (53 papers). Some miRs have only been associated with cancer EMT. For example, miR-22 has been associated with cancer EMT in 25 papers, but never with noncancer EMT. In total, 40 miRs have been associated with cancer EMT in more than three papers but never with noncancer EMT (Supplementary Table 1), although the specificity of association was statistically significant only in a few cases (probably because of the paucity of noncancer studies). Representative miRs are shown in Table 1.

Target prediction implicates miR-182 in uncoupling of RECK expression from EMT in cancer. Second, we screened for miRs potentially targeting *RECK* mRNA using four algorithms as follows: miRanda, TargetScan, PicTar, and DIANA-microT. These algorithms evaluate the following features in different combinations and emphases: (1) base pairing pattern between a miR and a possible target, (2) thermodynamic stability of miR-mRNA hybrid, and (3) conservation of target sequences across related species. Despite the differences in their theoretical bases and computational procedures,[24] the results from these four algorithms showed considerable overlap. Among the 16 miRs with an mirSVR score (miRanda) less than −1.2, 6 miRs were also detected with the other algorithms (Table 2). Of note, miR-182 gave the top score with miRanda (mirSVR) and high scores with Pictar (PicTar score), TargetScan (Context score and P_{CT}), and MicroT-CDS (miTG score; Table 2 and Supplementary Tables 2–5).

Table 1. Number of original research papers describing miRs in the context of cancer and/or noncancer EMT.

	miR	CANCER EMT	NONCANCER EMT	RATIO (CANCER/NONCANCER)	*P*-VALUE
(1)	miR-22	25 (0.028)	0 (0.000)		0.041
	miR-10b	11 (0.012)	0 (0.000)		0.178
	miR-222	11 (0.012)	0 (0.000)		0.178
	miR-143	9 (0.010)	0 (0.000)		0.223
(2)	miR-200c	121 (0.136)	10 (0.069)	1.978	0.024
	miR-200b	75 (0.085)	11 (0.076)	1.115	0.726
	miR-200a	59 (0.067)	9 (0.062)	1.072	0.841
	miR-21	53 (0.060)	20 (0.138)	0.433	0.001
	miR-141	34 (0.038)	6 (0.041)	0.926	0.860
	miR-204	8 (0.009)	3 (0.021)	0.436	0.205
(3)	miR-15a	2 (0.002)	4 (0.028)	0.082	0.000
	miR-193a	0 (0.000)	1 (0.007)	0	0.013
	miR-302b	0 (0.000)	1 (0.007)	0	0.013
	miR-184	0 (0.000)	2 (0.014)	0	0.000
(4)	miR-183	7 (0.008)	0 (0.000)		0.283
	miR-182	6 (0.007)	0 (0.000)		0.321
	miR-96	4 (0.005)	0 (0.000)		0.418
	miR-183-96-182	15 (0.017)	0 (0.000)		0.115
	Total	887	145		

Notes: The numbers in parentheses indicate their proportion to all papers describing cancer EMT (second column) or noncancer EMT (third column). Ratio: ratio between the proportions (the numbers in the parentheses). *P*-value: probability calculated using two-tailed prop.test for null hypothesis that [proportion in cancer EMT] = [proportion in noncancer EMT]. The top miRs in each of the following four categories are listed: (1) enriched in cancer EMT, (2) enriched in both cancer and noncancer EMT, (3) enriched in noncancer EMT, and (4) members of the miR-183-96-182 cluster (individually and in total). For full results, see Supplementary Table 1.

Table 2. *RECK*-targeting miRs predicted by four commonly used algorithms.

miR	PicTar SCORE	miTG SCORE	P_{CT}	CONTEXT SCORE	mirSVR SCORE
miR-182	2.46	0.987	0.72	−0.29	−1.3325
miR-429		1	0.9	−0.28	−1.323
miR-200b		0.999	0.9	−0.29	−1.3226
miR-200c	1.66	0.999	0.9	−0.29	−1.3187
miR-195	2.84	0.995	0.89	−0.34	−1.3113
miR-424		0.97	0.89	−0.36	−1.3107
miR-15b	2.84	0.997	0.89	−0.39	−1.3072
miR-15a	2.84	0.997	0.89	−0.38	−1.3072
miR-497		0.972	0.89	−0.38	−1.3065
miR-590		0.961	0.63	−0.34	−1.2763
miR-21	1.79	1	0.63	−0.35	−1.2763
miR-221		0.741	0.31	−0.33	−1.2621
miR-222		0.858	0.31	−0.34	−1.2588
miR-216a					−1.2371
miR-148b					−1.2024
miR-152					−1.2024
miR-96	2.46		0.73	−0.13	−0.5472
miR-183					

Notes: miRs with a mirSVR score smaller than −1.2 and members of miR-183-96-182 are shown. For full lists, see Supplementary Tables 2–5. Lower values predict higher probabilities of *RECK*-targeting in the case of mirSVR (miRanda) and Context score (TargetScan), while higher scores predict higher probabilities in the case of PicRar score, miTG score (MicroT-CDS), and P_{CT} (TargetScan).

Inverse correlation between the levels of miR-182 and *RECK* mRNA in breast cancer tissues. Third, we looked for miRs that show differential expression between cancer and normal tissues with correlation to *RECK* mRNA expression levels. For this purpose, we chose to use TCGA breast cancer data in which both mRNA and miR data in cancer tissues as well as matched normal tissues were available ($n = 53$). As expected, highly significant downregulation of *RECK* mRNA in cancer tissues was observed ($P = 0.000$; Fig. 2A). When we chose the cutoff values of 0.8 and 1.25 for downregulation and upregulation, respectively, we found that *RECK* mRNA was downregulated in cancer tissue in 47 patients, upregulated in two patients, and unchanged in four patients (Fig. 2B).

In the case of miRs, distribution of the median of the ratio (cancer/normal) was roughly symmetrical around zero (in logarithmic scale). When cutoff values of 0.5 and 2 were chosen, 71 miRs were downregulated and 77 miRs were upregulated in cancer tissues (Fig. 2C). For a technical reason (to avoid division by zero), we assigned the expression value of 0.1 to all undetectable miRs. Therefore, the miRs undetectable in both cancer and normal tissues would give the ratio of 1. Among the 638 miRs that gave the ratio of 1 (0 in the $\log_2$ scale as shown in Fig. 2C), 279 miRs were undetectable in all samples.

Based on these classifications, we compared the changes in the expression of *RECK* mRNA and miRs between cancer tissues and matched normal tissues using a newly developed nonparametric correlation test, named the PDC test (see "Methods" section for detail; the distribution of diff.sum based on simulation is illustrated in Fig. 2D). Through this method, we obtained a list of 1046 miRs, whose expression is similarly or inversely correlated with that of *RECK* mRNA (Table 3; for full list, see Supplementary Table 6). In these tables, "diff.sum" score represents the extent of inverse correlation between the levels of the miR and *RECK* mRNA. In the present comparison of 53 pairs of data, diff.sum scores, by definition, range between 0 (perfect positive correlation) and 106 (perfect inverse correlation). The diff.sum scores for miR-139 (9) and miR-486-1 (13) are among the lowest in the list (Table 3, see also Fig. 2E), indicating that the changes in their expression are concordant to that of *RECK* mRNA in most cases. Interestingly, both miRs have been reported as tumor suppressors, and their expression has been shown to be inversely correlated with tumor proliferation or progression.[25,26] On the other hand, the diff.sum scores for miR-96 (91) and miR-183 (95) are among the highest in the list, indicating their inverse correlation with *RECK* mRNA. miR-183, miR-96, and miR-182 are located in close proximity on human chromosome 7 and are expressed in a single primary transcript.[27] Indeed, miR-182 also shows a high diff.sum score (90). Of note, several well-studied miRs reported to exhibit oncogenic activities, such as miR-21, or those of bidirectional activities, such as the miR-200 family (miR- 200a, miR-200b, and miR-200c), are also among the high diff.sum score group (ie, inversely correlated with *RECK* mRNA).

TEAD4 and SOX2 as possible miR-183-96-182 regulators. The above findings prompted us to focus on the miR-183-96-182 cluster. Besides its inverse correlation in expression with *RECK* (Table 3), two of its members (miR-182 and miR-96) have the potential to target *RECK* (Table 2) and at least four studies have been published for each member (15 papers in total) in the context of cancer EMT, but never in the context of noncancer EMT (Table 1).

The TSS of the miR-183-96-182 primary transcript has been predicted to be at 5207 bp upstream of the miR-183 precursor on human chromosome 7.[28] A region of 4 kb surrounding this TSS is rich in nucleosomes with H3K4Me3 and H3K27Ac modifications, both known to mark active transcription. ENCODE ChIP-seq data also indicate binding of many (~70) transcription factors within the region of about 16 kb surrounding the TSS (Fig. 3A). A survey of gene expression databases indicated that several of these factors may affect the expression of *RECK* in various experimental settings. For instance, knockdown of TEAD4 resulted in elevated *Reck* expression in the mouse myogenic cell line C2C12 (GSE27845, Fig. 3B). Likewise, suppression of Ezh2 resulted in elevated *Reck* expression in high-grade glioma and preadipocytes (GSE63853 and GSE20054, respectively, Supplementary Fig. 5A and B). No miR-183-96-182 expression data

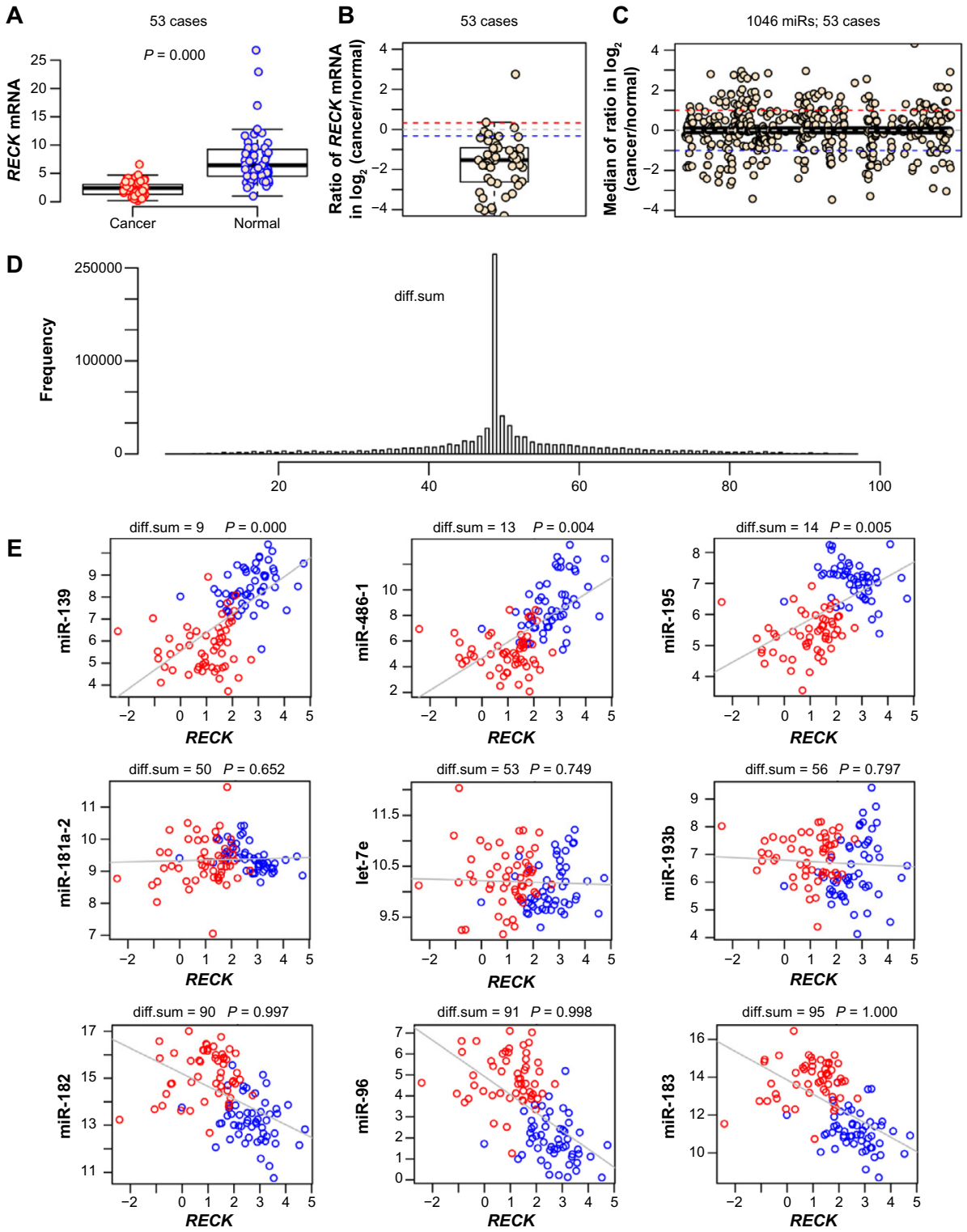

Figure 2. PDC test for *RECK* mRNA and various miRs in paired TCGA breast cancer samples. (**A**) Expression of *RECK* mRNA in 53 pairs of cancer and matched normal tissues in TCGA breast cancer dataset. The boxes indicate the interquartile range (IQR) of data between 75% (Q3) and 25% (Q1). The bars below and above each box indicate the data in Q1 − 1.5 × IQR and Q3 + 1.5 × IQR, respectively. *P*-value was calculated by paired *t*-test. (**B**) Ratio (in $\log_2$) of the *RECK* mRNA levels between cancer and normal tissues from 53 breast cancer patients. Blue and red broken lines indicate the cutoff values for "Down" (0.8) and "Up" (1.25) groups, respectively. (**C**) Median of ratios (in $\log_2$) of the levels of miRs (1046 species) among 53 pairs of cancer and normal tissues. Blue and red broken lines indicate the cutoff values for "Down" (0.5) and "Up" (2.0) groups, respectively. (**D**) Distribution of diff.sum based on random shuffling of original miRs diff matrix and then generating a 1000 times larger simulated dataset (see "Methods" section for details). (**E**) Relationship between the levels (in $\log_2$) of *RECK* mRNA and various miRs. Top row: miRs with low diff.sum scores (positive correlation). Middle row: miRs with intermediate diff.sum scores. Bottom row: miRs with high diff.sum scores (inverse correlation). Red spots represent cancer samples and blue spots represent normal tissue samples. Gray line represents regression curve. *P* represents the cumulative probability obtained from the distribution showed in (**D**), and values sufficiently close to 0 or 1 both indicate a rare event. For more completed scatter between top miRs with low/high diff.sum scores and *RECK* mRNA, see Supplementary Figures 2 and 3.

Table 3. miRs exhibiting positive or inverse correlation in expression with *RECK* mRNA among the matched breast cancer and normal breast tissues from 53 patients (TCGA data).

RANKING	miR	diff.sum	*P*	REFERENCES	SUMMARY
T1	miR-139	9	0.000	24204738, 24942287	Tumor suppressor
T2	miR-486-1	13	0.004	25104088, 21415212	Tumor suppressor
T3	miR-10b	14	0.005	21067538, 18948893	Context-dependent oncomiR or bidirectional?
T3	miR-195	14	0.005	24402230, 24787958	Tumor suppressor
T3	miR-204	14	0.005	25157435, 23204229	Tumor suppressor
T6	miR-451a	15	0.008	24918822, 24841638	Tumor suppressor
T6	miR-99a	15	0.008	24957100, 21878637	Tumor suppressor
T8	miR-140	16	0.009	23401231, 24971538	Tumor suppressor
T8	miR-584	16	0.009	21119662	Tumor suppressor
T10	miR-100	17	0.013	24586203	Bidirectional
T10	miR-1247	17	0.013	24785261	Tumor suppressor
T10	miR-125b-2	17	0.013	22711523	Tumor suppressor
T10	miR-133a1	17	0.013	23723074, 25198665	Tumor suppressor
T10	miR-144	17	0.013	25073510, 25961751	Tumor suppressor
	…	…	…	…	…
B41	miR-200c	76	0.963	25826661, 18376396	Bidirectional, targets *RECK*
B33	miR-200b	77	0.968	25826661, 18376396	Bidirectional, targets *RECK*
B33	miR-7-1	77	0.968	22761427, 25027403	Bidirectional, targets *RECK*
B8	miR-141	85	0.991	25003366, 25008569	Bidirectional
B8	miR-200a	85	0.991	23679328, 25239643	Bidirectional
B8	miR-301b	85	0.991	24398967	OncomiR
B8	miR-92b	85	0.991	24162673, 23416699	OncomiR, targets *RECK*
B7	miR-429	87	0.995	24866238, 24572141	OncomiR
B6	miR-1301	88	0.995	22159405	Possible tumor suppressor
B5	miR-182	90	0.997	23383207, 23333633	OncomiR, targets *RECK*
B3	miR-21	91	0.998	20447717, 25084400	OncomiR, targets *RECK*
B3	miR-96	91	0.998	24366472, 24469470	OncomiR, targets *RECK*
B2	miR-592	92	0.999		Unknown
B1	miR-183	95	1.000	23538390, 25337200	Bidirectional

Notes: Representative references are cited (PMID). Ranking: T, from top; B, from bottom. *P* represents the cumulative probability. For full list, see Supplementary Table 6.

are available from these experiments. Paradoxically, however, knockdown of EZH2 resulted in elevated expression of miR-183, miR-182, and miR-182* in the prostate cancer cell line DU145 (GSE26996, Supplementary Fig. 5C). No *RECK* expression data are available from this experiment.

We also surveyed the effects of additional 30 factors, mainly oncogenes and EMT regulators, on *RECK* and miR-183-96-182 expression. This survey detected SOX2, whose knockdown influenced the expression of the miR-183-96-182 cluster and *RECK*. In one dataset (GSE67993), SOX2 knockdown in human embryonic stem cells resulted in downregulation of miR-182/miR-96 and upregulation of *RECK* mRNA (Fig. 3C). In another dataset (GSE49403), knockdown of SOX2 or its coregulator *RMST* in neural stem cells resulted in upregulation of *RECK* mRNA (Fig. 3D), although this dataset did not provide information on miR-183-96-182 expression. In addition, another dataset (GSE59380) indicated that knockdown of SOX2 in the embryonal carcinoma cell line 2102Ep resulted in downregulation of miR-183/miR-182.[29]

Discussion

Utility of the PDC test developed in this study. One of the questions we attempted to address in this study was when the level of *RECK* mRNA changes (between cancer and normal tissues), how the levels of various miRs change in the same patient. The change in this case can be qualitative, since it involves intrinsic differences between cancer tissue and the corresponding normal tissue. Our PDC test is a simple approach to make such comparisons possible. In this test, we

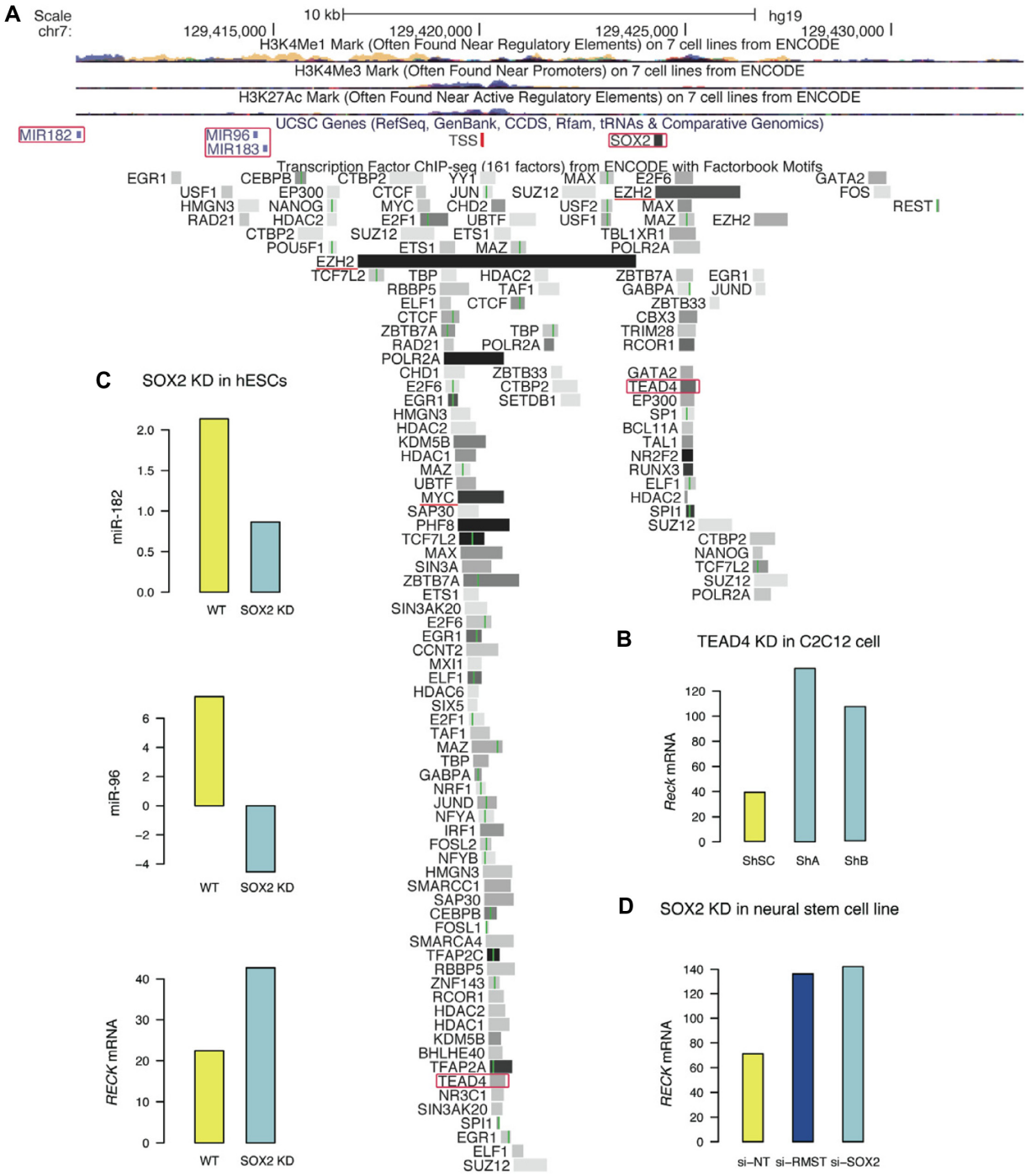

Figure 3. Candidates for the transcription factors regulating the miR-183-96-182 cluster. (**A**) The 23-kb human genomic region containing the miR-183-96-182 cluster as shown in the UCSC genome browser. Three miRs, SOX2, and TEAD4 are highlighted by red boxes. MYC and EZH2 are underlined. The binding site for SOX2 is based on Vencken's meta-analysis (PMID: 25156079). TSS of pri-miR-183-96-182 is based on Chien et al (PMID: 21821656). (**B**) Effects of TEAD4 knockdown (KD) on the level of *Reck* mRNA in differentiated C2C12 cells. ShSC, control shRNA; ShA and ShB, two independent shRNAs targeting TEAD4; $n = 1$. (**C**) Effects of SOX2 KD on the levels of miR-182, miR-96, and *RECK* mRNA in undifferentiated human embryonic stem cells. WT, treated with scrambled siRNA; SOX2 KD, treated with SOX2 siRNA; $n = 1$. (**D**) Effects of knocking down SOX2 or *RMST* on the level of *RECK* mRNA in neural stem cells. si-NT, control siRNA; si-RMST and si-SOX2, siRNA targeting *RMST* and SOX2, respectively; $n = 1$. (**B**), (**C**), and (**D**) are based on the following NCBI GEO datasets, respectively, GSE27845, GSE67993, and GSE49403.

can choose cutoff values depending on the nature of the data being compared. For example, the levels of *RECK* mRNA are known to be generally low among cancer samples, and therefore, a too strict cutoff value (eg, 0.5) for downregulation may cause unnecessary loss of information. In the case of miRs, however, more robust cutoff values of 0.5 should be more appropriate in light of their mechanism of action. However, even when we chose more stringent conditions (cutoff values

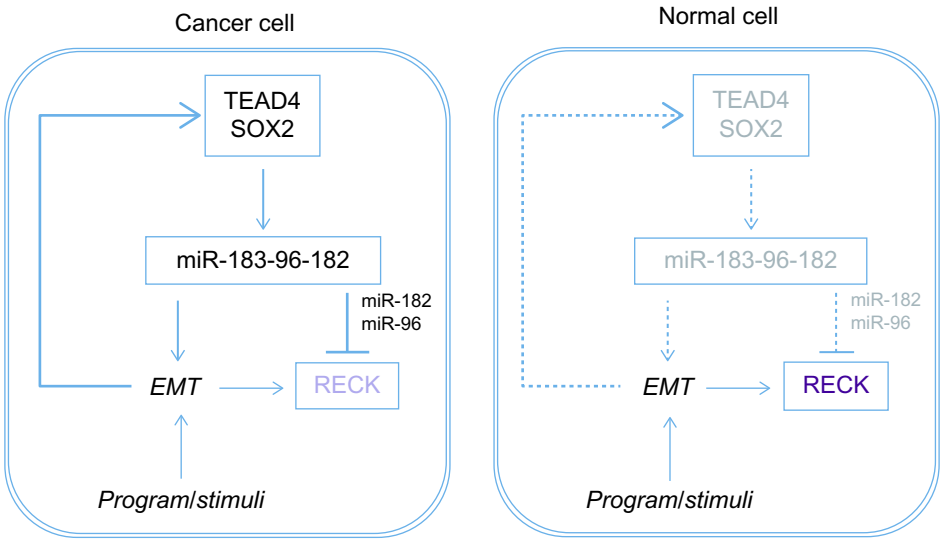

Figure 4. Model consistent with our findings. In cancer cells, transcription factors such as SOX2 and TEAD4 are upregulated and induce miR-182-96-183 expression. Three miRs in turn promote EMT, while two of them (miR-182 and miR-96) target *RECK* mRNA. The positive feedback loop between EMT and miR-182-96-183 works to stably downregulate RECK after EMT.

of 0.5 and 2 for both *RECK* mRNA and miRs), the top five miRs detected in this study were the same, albeit in different order, demonstrating the robustness of this method. Nevertheless, the adjustability of cutoff values should make this method useful in wider applications.

In this study, the PDC test was successfully used for enriching miRs, whose expression are positively or inversely correlated with *RECK* expression. Of note, among the top 14 miRs with positive correlation with *RECK* mRNA (diff.sum $\leq$17, $P \leq 0.013$; Table 3, Ranking T1–T10), 12 miRs have previously been reported to have tumor suppressor function (Table 3). Among the top 11 miRs with inverse correlation with *RECK* mRNA (diff.sum $\geq$85, $P \geq 0.991$; Table 3, Ranking B1–B8), 6 miRs have previously been reported as oncomiRs, and importantly, 4 of them (miR-96,[30–32] miR-21,[33–36] miR-182,[37–39] and miR-92b[40]) target *RECK* directly. Furthermore, additional three miRs reported to target *RECK* mRNA (miR-7-1,[36] miR-200b,[41] and miR-200c[41]) are also found within the top 41 in this category. Hence, this approach seems to be useful for predicting potential oncomiRs and tumor suppressor miRs and, in more general, for predicting potential miR-target relationship.

Classical theory holds that miRs repress target gene expression through two mechanisms as follows: (1) perfect or nearly perfect base pairing with the target mRNA promotes cleavage of the mRNA and (2) partially complementary miRs inhibit translation of target mRNAs. However, more recent findings indicate that translational inhibition by partially complementary miRs is also accompanied by accelerated degradation of target mRNA.[42] Our approach (to find miRs that target *RECK* and show inverse correlation in expression with *RECK*) is based on this new understanding.

miR-183-96-182 may uncouple RECK expression from EMT in cancer. We used three approaches to find candidate miRs that play roles in RECK suppression after EMT in cancer cells. The first approach of surveying published reports in PubMed has an advantage that these findings have been experimentally validated, but it also has disadvantages such as the large influence of key word choice and the arbitrary nature of the areas covered by current literature. The second approach of finding miR targets using multiple algorithms is more objective and genome wide (advantage), but the results are not always accompanied by experimental validation (disadvantage). The third approach of finding miRs with correlated changes in expression with a target gene is also objective and genome wide (advantage), but the results largely depend on the datasets used and do not usually provide any information on the mechanism of correlated changes (disadvantage). Although the results obtained through each approach may have limited value for identifying feasible candidates, the miRs detected by multiple approaches should be of particular interest.

The members of the miR-183-96-182 cluster were detected by all three approaches. First, these miRs are enriched in the abstracts related to cancer EMT studies (Table 1). Second, miR-182 was among the top-ranking miRs predicted to target *RECK* by all the algorithms used. miR-96 was also predicted to target *RECK*, albeit at lower ranking (Table 2). Third, miR-183, miR-96, and miR-182 were among the top-ranking miRs with inverse correlation with *RECK* in our PDC test (Table 3). Indeed, Hirata et al already showed that miR-182 could suppress RECK expression in cell lines derived from bladder cancer[38] and prostate cancer.[39] Multiple studies also indicate that miR-96 targets *RECK*.[30–32]

Several findings suggest the causative role of miR-183-96-182 in cancer EMT. Yang et al found that overexpression of miR-182 enhanced cell proliferation, invasion, and migration and induced molecular features of EMT, including

upregulation of Snail and Vimentin with concomitant downregulation of E-cadherin.[43] In a recent study, the three members of miR-183-96-182 cluster were found to promote EMT in breast cancer cells.[44] However, multiple lines of evidence place miR-183-96-182 downstream of EMT as well. In gallbladder cancer cells, TGFβ was found to induce miR-182 expression, and inhibition of miR-182 resulted in suppression of TGFβ-induced cancer cell migration and invasion.[45] Although TGFβ is not equivalent to EMT, it is commonly used to induce EMT in cultured cells. Likewise, glioma cells treated with TGFβ showed remarkable increase in the expression of miR-182, miR-183, and miR-96.[46] Taken together, these reports suggest a positive feedback loop between EMT and miR-183-96-182 in cancer cells. Interestingly, TGFβ showed no effect on miR-182 expression in normal human astrocytes,[46] supporting the idea that miR-182 may be involved in cancer-specific inhibition of RECK induction after EMT. Further strengthening this idea, miR-182 was downregulated after EMT in a prostate epithelial cell line (noncancerous).[47] However, this study was categorized into "cancer EMT" (Table 1) because the abstract contains the terms "cancer" and "metastasis", exemplifying a potential pitfall in such text-mining approaches.

miR-183-96-182 or its components are overexpressed in various cancers and may play positive roles in carcinogenesis.[43,44,48–51] Higher expression of the miR-183-96-182 cluster (as a whole or in part) in breast cancer cell lines compared with the MCF10A human breast epithelial cell line has also been reported in multiple studies.[27,32] However, there are several studies reporting a inverse correlation between miR-183-96-182 or its components and malignancy,[52,53] suggesting context-dependency of their actions.

Possible regulators of miR-183-96-182 expression. Several molecules such as β-catenin/TCF/LEF,[54] p53,[55] and TGFβ[46] have been implicated in the regulation of miR-183-96-182 expression.[48] In this study, however, we found some evidence suggesting SOX2 and TEAD4 as potential regulators of miR-183-96-182 expression.

SOX2 seems to have multiple effects on carcinogenesis: in lung cancer, SOX2 expression has been correlated with better prognosis,[56,57] while in esophageal cancer, SOX2 amplification and/or overexpression has been associated with poorer prognosis.[58] In breast cancer, high SOX2 expression has been associated with cancer formation and malignancy.[59–61] From three datasets on SOX2 knockdown experiments, we found evidence supporting the model that SOX2 suppresses RECK expression via upregulation of miR-183-96-182 in cancer cells (Fig. 4). Through a meta-analysis of two sets of SOX2 chromatin immunoprecipitation data using human embryonic stem cells, Vencken et al.[29] suggested a SOX2-binding site at ~4 kb upstream of the predicted miR-183 TSS (Fig. 3A); in that study, however, SOX2 was considered to be an inhibitor of EMT during embryonic development. In the context of cancer, however, SOX2 has been reported to bind to the promoters of *SNAIL*, *SLUG*, and *TWIST* and to promote EMT,[62] which fits with our model. Induction of Sox2 by TGF-β has also been reported.[63] Thus, although multiple lines of evidence in the databases support the involvement of SOX2 in the pathway proposed here (Fig. 4), the evidence is fragmentary and not always in comparable contexts.

TEAD4, a member of the TEAD family, plays crucial roles in mammalian development and carcinogenesis and is overexpressed in various types of cancer (Supplementary Fig. 4). Knockdown of TEAD4 in C2C12 cells resulted in marked upregulation of *Reck* expression (Fig. 3D). Based on ENCODE ChIP-seq data, the nearest TEAD4-binding site is quite distant in the case of *RECK* (27.8 kb upstream of TSS) but very close in the case of miR-183-96-182 (0.2 kb upstream of TSS) (Fig. 3A). In ovarian cancer, increased TEAD4 expression predicts poorer prognosis, and increased expression of both YAP and TEAD4 shows even more dramatic association with poor prognosis.[64] While this article was under revision, Liu et al published a study reporting TEAD4 overexpression in colorectal cancer and its activity to promote EMT by activating Vimentin expression.[65] Herein, we suggest that RECK inhibition via miR-183-96-182 activation may also be involved in TEAD4-induced malignant behaviors of cancer cells after EMT (Fig. 4).

Conclusions

Through multiple *in silico* approaches, we aimed to narrow down the candidate miRs possibly involved in cancer-specific suppression of EMT-induced RECK upregulation. Two oncomiRs (miR-182 and miR-96) closely clustered on human chromosome 7 show inverse correlation in expression with *RECK*, are capable of directly targeting *RECK*, and have previously been implicated both upstream and downstream of EMT in cancer cells. We also predict that transcription factors such as SOX2 and TEAD4 may enhance the expression of the miR-183-96-182 cluster from which miR-182 and miR-96 are produced (Fig. 4).

Acknowledgments

The authors thank all LMCR members, especially Drs. Masakazu Toi, Satoru Ikeda, Kazuki Matsui, Yoshiteru Murofushi, Shun Ikeda, Nobuhiro Okada, and Boban Stanojevic, for their support and valuable discussion; Drs. Ryo Yamada and J. B. Brown for their instructions on some of the statistical and mathematical methods; and Takeshi Yamasaki and Kaori Sugiyama for administrative assistance.

Author Contributions

Conceived and designed the experiments: ZW, MN. Analyzed the data: ZW, RM. Wrote the first draft of the manuscript: ZW. Contributed to the writing of the manuscript: MN, RM, KY, YY. Developed the structure and arguments for the paper: ZW, MN. Made critical revisions and approved the final version: MN. All the authors reviewed and approved the final manuscript.

Supplementary Materials

Supplementary Figure 1. Genomic location of 5 probes for *RECK* in AgilentG4502A_07.

Supplementary Figure 2. Scatter plot of top 36 miRs positively correlated with *RECK* in paired TCGA breast cancer samples.

Supplementary Figure 3. Scatter plot of top 36 miRs inversely correlated with *RECK* in paired TCGA breast cancer samples.

Supplementary Figure 4. ONCOMINE profile of transcription factors that might bind to miR-183-96-182 promoter.

Supplementary Figure 5. Effect of EZH2 suppression on *RECK* and miR-183-96-182 cluster expression.

Supplementary Table 1. miRs reported in cancer vs. non-cancer EMT studies.

Supplementary Table 2. miRanda prediction of miRs that target *RECK*.

Supplementary Table 3. TargetScan prediction of miRs that target *RECK*.

Supplementary Table 4. PicTar prediction of miRs that target *RECK*.

Supplementary Table 5. MicroT-CDS prediction of miRs that target *RECK*.

Supplementary Table 6. miRs that positively or inversely correlated with *RECK* based on analysis of TCGA paired breast cancer dataset.

REFERENCES

1. Takahashi C, Sheng Z, Horan TP, et al. Regulation of matrix metalloproteinase-9 and inhibition of tumor invasion by the membrane-anchored glycoprotein RECK. *Proc Natl Acad Sci U S A*. 1998;95(22):13221–6.
2. Noda M, Oh J, Takahashi R, Kondo S, Kitayama H, Takahashi C. RECK: a novel suppressor of malignancy linking oncogenic signaling to extracellular matrix remodeling. *Cancer Metastasis Rev*. 2003;22(2–3):167–75.
3. Clark JC, Thomas DM, Choong PF, Dass CR. RECK: a newly discovered inhibitor of metastasis with prognostic significance in multiple forms of cancer. *Cancer Metastasis Rev*. 2007;26(3-4):675–83.
4. Noda M, Takahashi C. Recklessness as a hallmark of aggressive cancer. *Cancer Sci*. 2007;98(11):1659–65.
5. Noda M, Takahashi C, Matsuzaki T, Kitayama H. What we learn from transformation suppressor genes: lessons from RECK. *Future Oncol*. 2010;6(7):1105–16.
6. Polyak K, Weinberg RA. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer*. 2009;9(4):265–73.
7. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014;15(3):178–96.
8. Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. *Annu Rev Pathol*. 2014;9:287–314.
9. Yuki K, Yoshida Y, Inagaki R, Hiai H, Noda M. E-cadherin-downregulation and RECK-upregulation are coupled in the non-malignant epithelial cell line MCF10A but not in multiple carcinoma-derived cell lines. *Sci Rep*. 2014;4:4568.
10. Rani J, Shah AB, Ramachandran S. pubmed.mineR: an R package with text-mining algorithms to analyse PubMed abstracts. *J Biosci*. 2015;40(4):671–82.
11. Gray KA, Yates B, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2015. *Nucleic Acids Res*. 2015;43(Database issue):D1079–85.
12. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 2008;36(Database issue):D149–53.
13. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol*. 2010;11(8):R90.
14. Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions. *Nat Genet*. 2005;37(5):495–500.
15. Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell*. 2007;27(1):91–105.
16. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 2009;19(1):92–105.
17. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol*. 2011;18(10):1139–46.
18. Paraskevopoulou MD, Georgakilas G, Kostoulas N, et al. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res*. 2013;41(Web Server issue):W169–73.
19. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
20. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006.
21. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC genome browser: year 5 update. *Nucleic Acids Res*. 2013;41(Database issue):D56–63.
22. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. 2007;23(14):1846–7.
23. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia*. 2004;6(1):1–6.
24. Witkos TM, Koscianska E, Krzyzosiak WJ. Practical aspects of microRNA target prediction. *Curr Mol Med*. 2011;11(2):93–109.
25. Liu R, Yang M, Meng Y, et al. Tumor-suppressive function of miR-139-5p in esophageal squamous cell carcinoma. *PLoS One*. 2013;8(10):e77068.
26. Oh HK, Tan AL, Das K, et al. Genomic loss of miR-486 regulates tumor progression and the OLFM4 antiapoptotic factor in gastric cancer. *Clin Cancer Res*. 2011;17(9):2657–67.
27. Li P, Sheng C, Huang L, et al. MiR-183/-96/-182 cluster is up-regulated in most breast cancers and increases cell proliferation and migration. *Breast Cancer Res*. 2014;16(6):473.
28. Chien CH, Sun YM, Chang WC, et al. Identifying transcriptional start sites of human microRNAs based on high-throughput sequencing data. *Nucleic Acids Res*. 2011;39(21):9345–56.
29. Vencken SF, Sethupathy P, Blackshields G, et al. An integrated analysis of the SOX2 microRNA response program in human pluripotent and nullipotent stem cell lines. *BMC Genomics*. 2014;15:711.
30. Guo H, Li Q, Li W, Zheng T, Zhao S, Liu Z. MiR-96 downregulates RECK to promote growth and motility of non-small cell lung cancer cells. *Mol Cell Biochem*. 2014;390(1–2):155–60.
31. Xia H, Chen S, Chen K, Huang H, Ma H. MiR-96 promotes proliferation and chemo- or radioresistance by down-regulating RECK in esophageal cancer. *Biomed Pharmacother*. 2014;68(8):951–8.
32. Zhang J, Kong X, Li J, et al. miR-96 promotes tumor proliferation and invasion by targeting RECK in breast cancer. *Oncol Rep*. 2014;31(3):1357–63.
33. Leite KR, Reis ST, Viana N, et al. Controlling RECK miR21 promotes tumor cell invasion and is related to biochemical recurrence in prostate cancer. *J Cancer*. 2015;6(3):292–301.
34. Zhang Z, Li Z, Gao C, et al. miR-21 plays a pivotal role in gastric cancer pathogenesis and progression. *Lab Invest*. 2008;88(12):1358–66.
35. Zhao W, Dong Y, Wu C, Ma Y, Jin Y, Ji Y. MiR-21 overexpression improves osteoporosis by targeting RECK. *Mol Cell Biochem*. 2015;405(1–2):125–33.
36. Jung HM, Phillips BL, Patel RS, et al. Keratinization-associated miR-7 and miR-21 regulate tumor suppressor reversion-inducing cysteine-rich protein with kazal motifs (RECK) in oral cancer. *J Biol Chem*. 2012;287(35):29261–72.
37. Chiang CH, Hou MF, Hung WC. Up-regulation of miR-182 by β-catenin in breast cancer increases tumorigenicity and invasiveness by targeting the matrix metalloproteinase inhibitor RECK. *Biochem Biophys Acta*. 2013;1830(4):3067–76.
38. Hirata H, Ueno K, Shahryari V, et al. Oncogenic miRNA-182-5p targets Smad4 and RECK in human bladder cancer. *PLoS One*. 2012;7(11):e51056.
39. Hirata H, Ueno K, Shahryari V, et al. MicroRNA-182-5p promotes cell invasion and proliferation by down regulating FOXF2, RECK and MTSS1 genes in human prostate cancer. *PLoS One*. 2013;8(1):e55502.
40. Lei L, Huang Y, Gong W. Inhibition of miR-92b suppresses nonsmall cell lung cancer cells growth and motility by targeting RECK. *Mol Cell Biochem*. 2014;387(1–2):171–6.
41. Pan Y, Liang H, Chen W, et al. microRNA-200b and microRNA-200c promote colorectal cancer cell proliferation via targeting the reversion-inducing cysteine-rich protein with Kazal motifs. *RNA Biol*. 2015;12(3):276–89.
42. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet*. 2015;16(7):421–33.
43. Yang MH, Yu J, Jiang DM, Li WL, Wang S, Ding YQ. microRNA-182 targets special AT-rich sequence-binding protein 2 to promote colorectal cancer proliferation and metastasis. *J Transl Med*. 2014;12:109.
44. Zhang W, Qian P, Zhang X, et al. Autocrine/paracrine human growth hormone-stimulated MicroRNA 96-182-183 cluster promotes epithelial-mesenchymal transition and invasion in breast cancer. *J Biol Chem*. 2015;290(22):13812–29.
45. Qiu Y, Luo X, Kan T, et al. TGF-β upregulates miR-182 expression to promote gallbladder cancer metastasis by targeting CADM1. *Mol Biosyst*. 2014;10(3):679–85.

46. Song L, Liu L, Wu Z, et al. TGF-β induces miR-182 to sustain NF-κB activation in glioma subsets. *J Clin Invest*. 2012;122(10):3563–78.
47. Qu Y, Li WC, Hellem MR, et al. MiR-182 and miR-203 induce mesenchymal to epithelial transition and self-sufficiency of growth signals via repressing SNAI2 in prostate cells. *Int J Cancer*. 2013;133(3):544–55.
48. Dambal S, Shah M, Mihelich B, Nonn L. The microRNA-183 cluster: the family that plays together stays together. *Nucleic Acids Res*. 2015;43(15):7173–88.
49. Ress AL, Stiegelbauer V, Winter E, et al. MiR-96-5p influences cellular growth and is associated with poor survival in colorectal cancer patients. *Mol Carcinog*. 2015;54(11):1442–50.
50. Lu YY, Zheng JY, Liu J, Huang CL, Zhang W, Zeng Y. miR-183 induces cell proliferation, migration, and invasion by regulating PDCD4 expression in the SW1990 pancreatic cancer cell line. *Biomed Pharmacother*. 2015;70:151–7.
51. Weeraratne SD, Amani V, Teider N, et al. Pleiotropic effects of miR-183~96~182 converge to regulate cell survival, proliferation and migration in medulloblastoma. *Acta Neuropathol*. 2012;123(4):539–52.
52. Li XL, Hara T, Choi Y, et al. A p21-ZEB1 complex inhibits epithelial-mesenchymal transition through the microRNA 183-96-182 cluster. *Mol Cell Biol*. 2014;34(3):533–50.
53. Kundu ST, Byers LA, Peng DH, et al. The miR-200 family and the miR-183~96~182 cluster target Foxf2 to inhibit invasion and metastasis in lung cancers. *Oncogene*. 2016;35(2):173–86.
54. Tang X, Zheng D, Hu P, et al. Glycogen synthase kinase 3 beta inhibits microRNA-183-96-182 cluster via the beta-Catenin/TCF/LEF-1 pathway in gastric cancer cells. *Nucleic Acids Res*. 2014;42(5):2988–98.
55. Chang CJ, Chao CH, Xia W, et al. p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol*. 2011; 13(3):317–23.
56. Wilbertz T, Wagner P, Petersen K, et al. SOX2 gene amplification and protein overexpression are associated with better outcome in squamous cell lung cancer. *Mod Pathol*. 2011;24(7):944–53.
57. Velcheti V, Schalper K, Yao X, et al. High SOX2 levels predict better outcome in non-small cell lung carcinomas. *PLoS One*. 2013;8(4):e61427.
58. Forghanifard MM, Ardalan Khales S, Javdani-Mallak A, Rad A, Farshchian M, Abbaszadegan MR. Stemness state regulators SALL4 and SOX2 are involved in progression and invasiveness of esophageal squamous cell carcinoma. *Med Oncol*. 2014;31(4):922.
59. Chen Y, Shi L, Zhang L, et al. The molecular mechanism governing the oncogenic potential of SOX2 in breast cancer. *J Biol Chem*. 2008;283(26):17969–78.
60. Lengerke C, Fehm T, Kurth R, et al. Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma. *BMC Cancer*. 2011;11:42.
61. Stolzenburg S, Rots MG, Beltran AS, et al. Targeted silencing of the oncogenic transcription factor SOX2 in breast cancer. *Nucleic Acids Res*. 2012;40(14):6725–40.
62. Herreros-Villanueva M, Zhang JS, Koenig A, et al. SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis*. 2013;2:e61.
63. Ikushima H, Todo T, Ino Y, Takahashi M, Miyazawa K, Miyazono K. Autocrine TGF-beta signaling maintains tumorigenicity of glioma-initiating cells through Sry-related HMG-box factors. *Cell Stem Cell*. 2009;5(5):504–14.
64. Xia Y, Chang T, Wang Y, et al. YAP promotes ovarian cancer cell tumorigenesis and is indicative of a poor prognosis for ovarian cancer patients. *PLoS One*. 2014;9(3):e91770.
65. Liu Y, Wang G, Yang Y, et al. Increased TEAD4 expression and nuclear localization in colorectal cancer promote epithelial-mesenchymal transition and metastasis in a YAP-independent manner. *Oncogene*. 2015. [Epub ahead of print]. doi: 10.1038/onc.2015.342